DOCUMENT RESUME

ED 365 686 TM 020 832

AUTHOR Young, Brian
TITLE Quadratic versus Linear Rules in Predictive Discriminant Analysis.
PUB DATE Nov 93
NOTE 22p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (22nd, New Orleans, LA, November 11-12, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classification; *Discriminant Analysis; *Equations (Mathematics); *Group Membership; Matrices; *Predictive Measurement; Research Methodology
IDENTIFIERS *Linear Models; Predictive Analysis; *Quadratic Equations; Statistical Analysis System; Statistical Package for the Social Sciences

ABSTRACT

Either linear or quadratic rules may be used to derive classification equations in discriminant analysis for the purpose of predicting group membership. Generally, the decision about which rule to use is governed by the degree to which the separate group covariance matrices are unequal. An example is presented that supports the superior internal classification hit rate of quadratic rules under conditions in which the sample matrices are unequal. The superiority of quadratic internal classification results provided by SAS relative to those provided by SPSS-X is also demonstrated. Finally, it is suggested that the potential external generalizability of the classification results also must be considered when deciding whether to use linear or quadratic rules to derive classification functions. Four tables. (Contains 16 references.) (Author)

Quadratic Versus Linear Rules in Predictive Discriminant Analysis

Brian Young
Department of Psychology
Texas A&M University 77843-4235

Running head: QUADRATIC VS. LINEAR RULES

Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, November 12, 1993.

Abstract

Either linear or quadratic rules may be used to derive classification equations in discriminant analysis for the purpose of predicting group membership. Generally, the decision about which rule to use is governed by the degree to which the separate group covariance matrices are unequal. An example is presented that supports the superior internal classification hit rate of quadratic rules under conditions in which the sample matrices are unequal. The superiority of quadratic internal classification results provided by SAS relative to those provided by SPSS-X is also demonstrated. Finally, it is suggested that the potential external generalizability of the classification results also must be considered when deciding whether to use linear or quadratic rules to derive classification functions.

Quadratic Versus Linear Rules in Predictive Discriminant Analysis

Classification functions, used for group prediction in discriminant analysis (DA), may be based on linear or quadratic rules. Essentially, the only procedural difference between these approaches lies in the covariance matrix that is chosen to derive the classification functions (prediction equations). If the covariance matrix is derived by pooling across the groups, the resulting functions are called linear; the functions are quadratic if the separate covariance matrices are used for the derivation of the prediction equations.

In general, DA assumes multivariate normality of the data, and equal covariance structures across groups if a linear rule is used. This latter assumption of the equality of group dispersion matrices should be tested, using Box's M or a Chi-square provided by standard statistical packages, before the equality of group means is tested. If the assumptions of multivariate normality and equal covariance structures across groups are met, a linear classification rule may be employed with relative confidence. However, even assuming multivariate normality, if the condition of equal covariance structures across groups is violated, a quadratic classification rule is often suggested as the more appropriate alternative.

Effects of Assumption Violations

There has been some debate about the effects of multivariate

normality assumption violations on DA results. This assumption is important for both tests of significance and for classification based on probabilities of group membership (Klecka, 1980, p. 61). It has been argued that deviations from normality may affect the results of quadratic classification much more than those of linear classification (Anderson & Bahadur, 1962; Johnson & Wichern, 1982). However, Eisenbeis and Avery (1972, p. 37) suggested that data that are not multivariate normal may be used in DA without biasing results to a noteworthy degree.

Inequality of the group dispersion matrices can have implications for tests of the equality of group means. Violation of this assumption results in a bias toward acceptance of the null hypothesis, which increases with the number of variables and the degree of inequality of the dispersions, and decreases in sample size (Holloway & Dunn, 1967). This is the case for both the chi-square test and F-test, because both are based on the formation of Wilk's lambda assuming multivariate normal populations with equal dispersion matrices. When the group means are close enough together so that the groups overlap, the differences between linear and quadratic classifications are particularly important; and these differences are likely to increase with the number of groups and the degree of group overlap (Eisenbeis & Avery, 1974). Since the power of the tests of group differences may be very low when the covariance matrices are different, it is important to test for the equality of the

dispersion matrices before testing for the equality of group means.

Why Quadratic Rules May Improve Classification

With fixed distances between group means, as the differences between group covariance matrices increases, the relative predictive ability of linear rules decreases, as compared to quadratic rules. This happens because linear rules use a common within-groups matrix that is computed by pooling the separate group covariance matrices. As a result, any function derived in this fashion will bias the results away from classification into groups with smaller variances in favor of those groups whose dispersions are larger (Eisenbeis & Avery, 1974). Klecka (1980, p. 61) notes that when group covariance matrices are unequal, the use of linear rules can result in distorted classification equations that do not provide maximum separation among groups, thereby distorting the probabilities of group membership.

With fixed distances between group variable means, as the number of variables increases, so will the discriminatory power of quadratic rules (Van Ness, 1979). This occurs because quadratic rules take advantage of the information provided by the different group covariances when making group classifications. That is, there are more variables with variance discrepancies to be utilized in deriving the classifications (Gilbert, 1969). This is particularly important in situations in which there is little distance between group variable means because, in these situations, the group variances provide most of the

discriminatory information, and provide all of it when group means are identical (Van Ness, 1979).

However, Van Ness (1979) found that the quadratic functions will begin to lose power as the number of variables increases to very large numbers, even with normal distributions and unequal covariance matrices, because sample covariance matrices become unreliable when there are too few research participants per variable. Huberty and Blommers (1974) also suggested that, with the existence of systematic biases toward or against groups, larger sample sizes will result in less accuracy when separate group covariances are used.

When Quadratic Rules May Impair Classification

Classification results often are determined "internally," meaning that objects or people are classified according to rules that were developed based on those objects or people. That is, internal classification results use existing information about group membership and variable scores to develop classification functions, and then test these same functions for accuracy by using them to reclassify the same individuals into groups. Internally developed classification functions based on quadratic rules are likely to result in a higher number of misclassifications when applied to subsequent samples than those based on linear rules, especially for small samples (Huberty & Wisenbacker, 1992; Michaelis, 1973). The more sample-specific information we use in prediction, the more accurate our sample-specific classification will be, but it is less likely that so

many features of our sample data will be replicated, resulting in less predictive accuracy of the same predictive equations in future samples. Michaelis (1973) found that, in most samples external quadratic classification gave better results than linear classification even with smaller sample sizes; however, the differences between internal and external DA are greatest with smaller sample sizes. Huberty (1975) discusses a study in which he found that in a "comparison of rules based on linear and quadratic equations using seven different data sets, internally, the quadratic rule was superior for all seven examples. However, the linear rule did as well if not better than the quadratic rule in an external sense" (p. 572).

In general, an important ultimate goal of science is the generalizability of theory and results across samples, settings, and time. Results derived using linear rules enjoy a number of advantages over those developed using quadratic rules in terms of generalizability to future samples. In general, the relative parsimony of linear classification functions is its greatest asset for generalizability. Because the quadratic rules use separate group covariance matrices, as contrasted from the one pooled-covariance matrix used by linear rules, there are more parameters to be estimated, and thus more opportunities for differences to occur from sample to sample. A related advantage of linear rules is that the use of one pooled covariance matrix results in the utilization of less sample-specific information, such as the variances of a particular group, thereby reducing

internal classification hit rates, but also decreasing the likelihood of external misclassifications.

A powerful technique for improving the generalizability of internally derived classification functions is the jackknife procedure. The leave-one-out (L-O-O) jackknife method classifies each subject based on a classification rule derived excluding that particular subject while using all of the remaining subjects. Unfortunately, none of the current statistical packages provides the L-O-O technique for the derivation of quadratic classification results (Huberty & Wisenbaker, 1992), thereby further limiting their potential generalizability.

Linear Vs. Quadratic Rules: A Heuristic Comparison of Internal Classification Results

Studies comparing quadratic and linear rules in the development of internal classification functions for different data sets have found quadratic rules to be superior (see Huberty, 1975), or as good if not better (Eisenbeis & Avery, 1974) than linear rules. Eisenbeis and Avery (1972) illustrate an example in which the overall performance (classification hit rate) for both rules was fairly comparable (in fact, the linear rule did slightly better); however, there were differences in the hit rates for particular groups. That is, the linear rule correctly classified more of the good loans for a bank whereas the quadratic rule correctly classified more of the bad loans. These differences would have potentially important consequences for the

bank because the identification of bad loans is far more critical than identifying good loans.

For illustrative purposes, the current paper presents comparisons of linear and quadratic classification results using the Sesame Street data base (Stevens, 1992, p. 578). These data were collected in a study conducted by the Educational Testing Service designed to see whether the television program taught preschool-related skills to members of five populations: 1) three to five year old inner-city disadvantaged children, 2) four year old advantaged suburban children, 3) advantaged rural children, 4) disadvantaged rural children, and 5) disadvantaged Spanish speaking children. The current analyses used the sampling site as the class variable, and the dependent variables were difference scores computed by subtracting the posttest from pretest scores on the various tests of knowledge of body parts, letters, forms, numbers, relations, and classification skills and scores on the Peabody Picture Vocabulary Test.

In addition, both SPSS-X and SAS statistical programs were used to analyze the data so that the results provided by each could be compared. This comparison is important because it has been argued that SAS is the only major statistical software package that provides accurate internal quadratic classification results (Huberty & Wisenbaker, 1992). The basic commands and subcommands used to perform the predictive DA's and develop the classification functions in both SAS and SPSS-X are provided in Table 1.

Insert Table 1 about here

In SAS, the rule used to derive the classification functions is evoked after "POOL =" and is indicated by either NO, TEST, or YES. When POOL = NO is chosen, the individual group covariance matrices (quadratic rule) are used to classify cases into groups. The default, POOL = YES, uses the pooled covariance matrix to compute linear functions from which cases are classified. POOL = TEST provides a statistical significance test of the homogeneity of the within group covariance matrices using Bartlett's likelihood ratio. This test is unbiased but not robust to non-normality (SAS, 1990). To use this, the option METHOD = NORMAL must be used rather than METHOD = NPAR (which is the default). The program then uses either a quadratic or linear rule to compute the classification functions depending on the outcome of the test.

In SPSS-X, the rule to be used in deriving the classification functions is evoked by the subcommand "CLASSIFY =" and is indicated by POOLED or SEPARATE. POOLED, which is the default, calls for the program to use the pooled covariance matrix to compute the linear functions for group classification. The SEPARATE routine uses the individual within-group covariance matrices for classification but does not provide mathematically correct quadratic results because the cases are classified based on the discriminant functions and not the observed variables (Huberty & Wisenbaker, 1992; SPSS, 1988; Tatsuoka, 1971).

Results

The results of the chi-square provided by SAS (using POOL = TEST) and Box's M provided by SPSS-X are both significant (chi-square (112 DF) = 171.97, $p < .05$; Box's M (40 DF) = 60.71, $p < .05$), indicating that, for these data, the within-group covariances are not constant across groups and thus, that quadratic rules may be appropriate. Classification functions were derived using quadratic and linear rules in both SPSS-X and SAS. The hit rates using linear rules for both statistical packages, the quadratic results from SPSS-X, and the quadratic results provided by SAS may be found in Tables 2, 3, and 4 respectively.

Insert Tables 2, 3, and 4 about here

As reflected by the overall percentages of correct group classification in each of these tables, the hit rates improved with the use of quadratic rules. SAS and SPSS-X provide identical internal linear classification results; however, as could have been expected, the quadratic classification results provided by SAS were different from those given by SPSS-X. Since only SAS yields mathematically correct internal quadratic classification results (Huberty & Wisenbaker, 1992), it was not surprising that the overall hit rates using quadratic rules in the SAS analyses were higher than the quadratic results provided by the SPSS-X analyses. As also may be observed in the tables, the number and percent of cases classified into each group

(whether correctly or incorrectly) changed according to whether linear rules, SAS quadratic rules, or SPSS-X quadratic rules were used.

The only group for which quadratic rules resulted in slightly fewer correct classifications was Group 1, and there was one less correct classification in SAS than in SPSS-X. Quadratic rules improved the classification rates for the other four groups and the quadratic results provided by SAS improved the classification rates for Groups 2, 3, and 5 beyond the improvements provided by the SPSS-X quadratic results.

Discussion

The presently reported results further support the superiority of internal quadratic classification results and are concurrent with other studies which have found them to be superior (see Huberty, 1975), or as good if not better (Eisenbeis & Avery, 1974) than internal linear classification results. The analyses also exemplify the differences between the internal quadratic classification results provided by SAS and SPSS-X, reflecting the mathematical inaccuracy of the SPSS-X results (Huberty & Wisenbaker, 1992) and the improved hit rates from quadratic classification using SAS. Since Sesame Street was specifically targeted to help disadvantaged children, the improved internal classification rates for Groups 4 and 5 using quadratic rules (especially for Group 5 using SAS) and only slight decrement for identifying Group 1 members, provide further evidence of the technique's power. Also, it should be noted

that, regardless of hit rates, many individual cases were classified into different groups depending on which rules were used, including whether quadratic classification was performed in SAS or SPSS-X. Thus, the choice of technique can have a large impact on the internal classification results.

When group covariance matrices are unequal, as in this example, quadratic rules result in better internal classification than linear rules because they utilize the extra information provided by the differences among the group covariance matrices. Since this information is augmented as more variables are added to the analyses, the differences between the quadratic and linear internal classification results would likely be even greater if more variables were added to these analyses, as explained previously.

However, although the analysis of this information results in the superiority of quadratic rules in an internal sense, it also reduces their parsimony by increasing the number of parameters to be estimated, and thus greatly impairs their external generalizability. Therefore, if the researcher's ultimate goal is to use DA to classify members of future samples into groups, it is recommended that a linear rule be applied in the development of the classification functions. If a quadratic rule is to be used in such situations, and as long as the L-O-O jackknife method is unavailable for the derivation of quadratic functions, then it is recommended that the classification functions' external hit rate be estimated using a holdout sample

initially excluded during the derivation of the functions. Future efforts should focus on improving the external validity of quadratic classification results. One possible direction for improvement would be the development of L-O-O jackknife methods for the development of internal quadratic classification functions in major statistical packages.

References

Anderson, T.W., & Bahadur, R.R. (1962). Classification into two multivariate normal distributions with different covariance matrices. Annals of Mathematical Statistics, 33, 420-431.

Eisenbeis, R.A., & Avery, R.B. (1972). Discriminant analysis and classification procedures: Theory and applications. Toronto: D.C. Heath.

Eisenbeis, R.A. & Avery, R.B. (1974). Comparison of linear and quadratic classification procedures with unequal sample dispersions. The Indian Economic Journal, 21, 95-104.

Gilbert, E.S. (1969). The effect of unequal variance-covariance matrices on Fisher's linear discriminant function. Biometrics, 505-515.

Holloway, L.N. & Dunn, O.J. (1967). The robustness of Hotelling's T^2, Journal of the American Statistical Association, 62 124-136.

Huberty, C.J. (1975). Discriminant analysis. Review of Educational Research, 45, 543-598.

Huberty, C.J., & Blommers, P.J. (1974). An empirical comparison of the accuracy of selected multivariate classification

rules. Multivariate Behavioral Research, 9, 59-84.

Huberty, C.J., & Wisenbaker, J.M. (1992). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in Social Science Methodology (Vol. 2, pp. 169-208). Greenwhich, CT: JAI Press Inc.
Press Inc.

Johnson, N., & Wichern, D. (1982). Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Klecka, W.R. (1980). Discriminant analysis. Newbury Park, CA: SAGE.

Michaelis, J. (1973). Simulation experiments with multiple group linear and quadratic discriminant analysis. In T. Cacoullos (Ed.), Discriminant analysis and applications.
New York: Academic Press.

SAS Institute Inc. (1990). SAS/STAT User's Guide (Version 6, 4th ed., Vol. 1). Cary, NC: SAS Institute Inc.

SPSS, Inc., (1988). SPSS-X User's Guide (3rd ed.). Chicago, IL: SPSS, Inc.

Stevens, J. (1992). Applied multivariate statistics for the

social sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Tatsuoka, M.M. (1971). Multivariate analysis. New York: Wiley.

Van Ness, J. (1979). On the effects of dimension in discriminant analysis for unequal covariance populations. Technometrics, 21, 119-127.

Table 1

Basic Commands and Subcommands Used to Obtain Classification Results

SAS Commands and Subcommands

```
PROC DISCRIM POOL = NO|TEST|YES <other options>;
     CLASS <group classification variable>;
```

SPSS-X Commands and Subcommands

```
DISCRIMINANT GROUPS = <group class. variable (low,high value)>
     /CLASSIFY = POOLED|SEPARATE
```

Table 2

SPSS-X and SAS Internal Linear Classification Results

Actual group	n	Number and percent classified into each group 1	2	3	4	5
Group 1	60	15 25.0%	10 16.7%	11 18.3%	13 21.7%	11 18.3%
Group 2	55	9 16.4%	32 58.2%	0 0.0%	6 10.9%	8 14.5%
Group 3	64	8 12.5%	4 6.3%	24 37.5%	18 28.1%	10 15.6%
Group 4	43	8 18.6%	6 14.0%	10 23.3%	15 34.9%	4 9.3%
Group 5	18	1 5.6%	4 22.2%	2 11.1%	3 16.7%	8 44.4%

Percent of correct group classifications: 39.17%

Table 3

SPSS-X Internal Quadratic Classification Results

Actual group	n	Number and percent classified into each group 1	2	3	4	5
Group 1	60	14(-1) 23.3%	9(-1) 15.0%	12(+1) 20.0%	16(+3) 26.7%	9(-2) 15.0%
Group 2	55	4(-5) 7.3%	34(+2) 61.8%	3(+3) 5.5%	3(-3) 5.5%	11(+3) 20.0%
Group 3	64	5(-4) 7.8%	3(-1) 4.7%	30(+6) 46.9%	14(-4) 21.9%	12(+2) 18.8%
Group 4	43	2(-6) 4.7%	3(-3) 7.0%	8(-2) 18.6%	25(+10) 58.1%	5(+1) 11.6%
Group 5	18	0(-1) 0.0%	4(+0) 22.2%	0(-2) 0.0%	4(+1) 22.2%	10(+2) 55.6%

Percent of correct group classifications: 47.08%

Note. The values in parentheses represent the change in number of cases classified to each group when the quadratic rule is used instead of the linear rule.

Table 4

SAS Internal Quadratic Classification Results

Actual group	n	Number and percent classified into each group 1	2	3	4	5
Group 1	60	13(-2) 21.7%	9(-1) 15.0%	17(+6) 28.3%	15(+2) 25.0%	6(-5) 10.0%
Group 2	55	5(-4) 9.1%	40(+8) 72.7%	4(+4) 7.3%	2(-4) 3.6%	4(-4) 7.3%
Group 3	64	5(-3) 7.8%	4(+0) 6.25%	36(+12) 56.25%	8(-10) 12.5%	11(+1) 17.2%
Group 4	43	2(-6) 4.65%	2(-4) 4.65%	7(-3) 16.3%	25(+10) 58.1%	7(+3) 16.3%
Group 5	18	0(-1) 0.0%	0(-4) 0.0%	0(-2) 0.0%	1(-2) 5.6%	17(+9) 94.5%

Percent of correct group classifications: 54.58%

Note. The values in parentheses represent the change in number of cases classified to each group when the quadratic rule is used instead of the linear rule.